真人与机器人交互研究的现状与展望——浅论心理学与人工智能的交叉*

邓兆鑫[1,2]　付超[1,2]　杨雪[1,2]　王益文[1,2]

[1]（福州大学心理与认知科学研究所 福州 350116）

[2]（福州大学中国社会信任研究中心 福州 350116）

摘要 机器人正逐步融入我们的社会，以往研究大多聚焦于机器人的软件编程或机械制造，鲜有研究着眼于心理学在人与机器人交互研究，特别是人与机器人二元关系建立、发展和变化中的重要性。鉴于社交机器人具有高于工业机器人的社交性，人与机器人交互研究中的机器人一般特指社交机器人。在介绍社交机器人的主要应用基础上，本文分类归纳出人与机器人交互中的信任、共情和社会距离等研究主题，阐述了研究中所涉及的恐怖谷、媒体等同和心智感知理论，探讨了社交机器人可能引发的伦理问题并对未来人与机器人交互的信任、共情等研究主题的深入和扩展进行了针对性地展望。心理学能够提供人类情感、认知和行为等方面的知识，帮助建立和形塑良好的人与机器人二元关系。

关键词 人与机器人交互；社交机器人；人工智能心理学

1　人与机器人交互研究的起源

科幻电影中描绘的机器人通常具有和人类相似的外观和思想，它们能够像人类一样体验情感、表达观念并采取行动。这样的场景距现实生活有多远？未来机器人是否真的可以像人类一样思考和感受？

机器人的制造历史可追溯到 1459 年。Lenonardo da Vinci 于此年设计出一个机械骑士，它由轮轴操纵，能移动手臂、升起面甲(Rosheim, 2006)。20 世纪中叶，Alan Turing 等人在机器思维方面的贡献，为现代数字计算和自主机器人(autonomous robot，不需要持续的人工引导就能自主完成任务的机器)的研发打下了坚实的基础(Turing, 1950)。20 世纪 50 年代，应用于汽车工业，可通过编程进行数字化操纵的机械臂是第一代真正意义上的机器人。此后，工业机器人不断发展并被逐渐应用到军事领域，如监听、排爆和自动化武器等方面。但无论是工业机器人还是军事机器人，它们多被人类视为工具而非工作伙伴或战友，人类与这些机器人之间并未形成一般意义上的交互。在心理学中，交互发生在两个或多个存在相互影响和相互作用的社会实体之间。因此，人与机器人形成交互的前提是机器人被人视作伙伴而非工具。人们如何看待机器人很大程度上取决于机器人的功能。其功能可分为实用功能和社会功能两类，实用功能体现在解放人力劳动、提高生产效率等方面，社会功能则体现在提供社会支持和陪伴交流等方面。传统而言机器人主要发挥其实用功能，但随着近 20 年来科技的发展，机器人的功能也逐步变化并向社会功能方面倾斜(Wang & Krumhuber, 2018)。机器人渐渐融入人们的日常生活，并在家庭、学校、医院等环境中成为人们的得力助手。它们(社

---

* 本文系国家自然科学面上基金项目（31771238）的研究成果之一，并受福建省闽江学者计划资助。

通讯作者：王益文，E-mail:wangeven@126.com

交机器人, social robot)所充当的角色不再仅仅是工具，而是人类的亲密伙伴。同时一些问题也引起了人们的思考，如：人和机器人协作效率的影响因素以及人类是否对机器人负有责任等。围绕这些问题，人与机器人交互（human-robot interaction, HRI）研究就此逐步展开。鉴于社交机器人所具有的高社交性，本文中人与机器人交互主要指人和社交机器人的交互。下文对社交机器人的应用进行简单梳理，并对人与机器人交互研究的现状及发展方向进行阐述。

2 社交机器人的主要应用

社交机器人正逐渐步入人们的生活，在一些养老院中，可以看到用于老年人日常陪护的医疗机器人；在部分教育机构中，可以看到用于儿童教育的教育机器人；此外，在一些大型购物中心，人们还能看到用于提供指引信息的向导机器人。科研人员试图制造出形态功能各异的机器人以满足不同使用群体的需求。本节对目前社交机器人的应用情况进行简要梳理。

2.1 医疗机器人

医疗机器人的应用主要面向两类群体：老年群体和儿童群体。

世界正面临人口老龄化危机，尤其是发达国家，医护人员已无法满足老年群体日益增长的医疗保健需求，越来越多的人试图通过机器人辅助来解决这一难题(Mann, MacDonald, Kuo, Li, & Broadbent, 2015)。机器人能帮助人们进行血压监测，可以给健忘者提供提醒服务，还可以辅助行动不便者行走、沐浴等。此外，在心理健康方面，机器人还能提供陪伴和交流。最著名的陪伴型社交机器人是源自日本的 Paro，其外形与格陵兰海豹幼崽相近(见图 1-a)，当传感器探测到触碰，光线、声音或位置变化时，Paro 便会发出海豹的叫声。有研究表明用于照看老年人的机器人外观要与它们所完成的任务相匹配，相对于机械化的机器人，人们更偏好毛绒绒的机器人进行陪伴(Broadbent, Tamagawa, Kerse, & Knock, 2009)。因此设计师选择了海豹幼崽的形象，他们希望人们能够像对待宠物狗一样，拥抱、爱抚 Paro 并与之进行交流。Paro 的主要使用场所在疗养院，它可像真人一样对老年人进行陪伴并在治疗老年痴呆症方面发挥作用，研究者对身处养老院的老年被试进行的随机控制实验表明，相对于真实的宠物狗，Paro 能够显著降低被试的孤独感并增加其社会互动行为，如更多地和机器人或他人进行交流(Robinson, Macdonald, Kerse, & Broadbent, 2013)。

面向儿童方面，有研究指出儿童的行为模式与成人有所区别，在和机器人交互的过程中，儿童更倾向于把机器人看做是有生命的实体，因而儿童和机器人的交互在人与机器人交互研究中占有特殊地位(Vallèsperis, Angulo, & Domènech, 2018)。该研究通过分析医疗情境下儿童对与社交机器人交互的想象来探索医疗机器人所应具备的特征。在实际应用中，医疗社交机器人可通过诱发儿童做出一些行为，并据此对其自闭程度进行评估以实现对自闭症患儿的诊疗，同时还能向儿童传授生活技能，增加儿童的亲社会行为等(Diehl, Schmitt, Villano, & Crowell, 2012)。自闭症患儿在理解他人意图及与他人交流方面存在障碍。相对真人而言，机器人的状态更易预测，活动范围也更小，自闭症患儿与其交互会更容易。通过和机器人交互，自闭症患儿能够习得社会规范并将之应用于真实的人际交互中。由于被试个体差异大、样本量不足且缺乏必要的控制组，目前机器人用于自闭症患儿诊疗的效果还需进一步提高(Diehl et al., 2012)。未来研究需要临床医生与心理学家的介入，用更好的实验设计来进一步提高此类机器人辅助疗法的效果。

2.2 教育机器人

教育机器人在教学领域中的应用分为两类：一是可自由组合并支持可视化编程的机器人(如乐高公司生产的 mindstroms，见图 1-b)，它们可充当教具，也可以帮助教师进行知识传授。二是交互式人形机器人(humanoid)，可直接作为教师来教授学生外语等课程(Mubin, Stevens, Shahid, Mahmud, & Dong, 2013)。有研究考察了机器人和真人教师教授儿童外语时教学效果的区别，结果发现面对机器人教师，儿童能更为主动地用外语交流，面对真人教师时则会显得羞涩犹豫。此外，机器人能够连续工作，这也克服了真人教师可能产生的疲劳问题(Chang, Lee, Chao, Wang, & Chen, 2010)。另有研究考察了此类机器人对低收入家庭儿童的教学效果，结果表明，相对于真人教师，机器人教师更能提高儿童的积极性，并且还能增强其社区意识和自我表达能力(Han, Park, & Park, 2015)。

2.3 向导机器人

在商场和博物馆等大型公共场所中人们常会见到向导机器人。除吸引顾客外，向导机器人还可提供场馆信息介绍，发送传单等服务。Sabelli 和 Kand (2015) 对日本一家配有向导机器人商场中的顾客进行的访谈结果表明，人们更多地把机器人看作是商场的吉祥物而不是一项公共设施，顾客喜欢机器人的存在并对在商城中放置机器人向导表示支持。另一项访谈研究表明，大多数顾客认为在商场里使用提供指引和发放传单服务的向导机器人是有益的，向导机器人不会像人类一样“以貌取人”，65%的受访者希望由机器人而不是人类提供同样的服务，超过90%的受访者表达了再次使用向导机器人的意愿(Satake, Hayashi, Nakatani, & Kanda, 2015)。

3. 人与机器人交互的研究主题

类人机器人(human-like robot)和类宠机器人(pet-like robot) 是人与机器人交互研究中主要关注的两类机器人。

类人机器人是指那些在外表和行为方面都高度接近人类的机器人。制造此类机器人是为了使人类能够依靠本能，更自然地和机器人交互。由中国科学技术大学研发的第三代特有体验交互机器人“佳佳”是国内首台类人社交机器人。它诞生于 2016 年 4 月，身高 1.6 米，具有精致的五官和接近真人的皮肤(见图 1-c)，具备人机对话理解、面部微表情展示、口型及躯体动作匹配等功能。研究团队认为机器人的形象与其品格和功能应协调一致，他们首次提出并探索了机器人品格的定义并赋予“佳佳”善良、勤恳、智慧等品格。在国外，高仿真类人机器人的研究同样吸引着科研人员。日本机器人学家 Ishiguro 所建造的孪生机器人(见图 1-d)是根据现实中的真人而制作的外观与其完全一致的机器人。通过远程操控，这些机器人能够表现出和真人一致的行为、交谈能力甚至性格。研究者认为，此类机器人能够反映出特定的真人形象，并帮助研究者更深入地理解认识人类的本质(Ishiguro & Nishio, 2007)。此外，美国布朗大学的 Phillips 研究团队研究了类人机器人的拟人程度。他们认为，类人机器人的外观设计大多依靠开发者的直觉，缺少对类人机器人范围、种类及组成特征之间关系的系统理解，据此，研究团队开发出在线测评工具以帮助研究人员对类人机器人的拟人程度进行评定(Phillips, Zhao, Ullman, & Malle, 2018)。

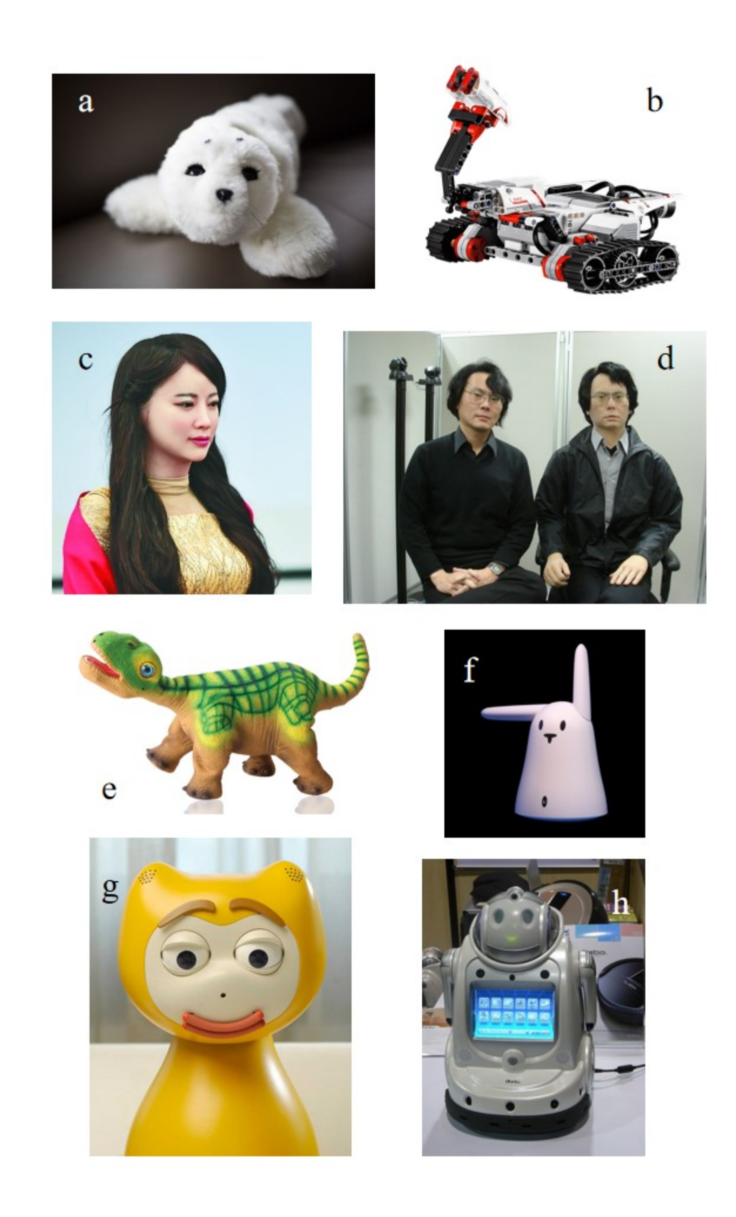


图 1 用于交互研究的机器人

类宠机器人是指外观接近小动物或宠物的机器人。上文所提日本产业技术研究所研发的 Paro 以及由美国 UGOBE 公司研发的 Pleo 都是典型的类宠机器人。Pleo 具有近似恐龙幼崽的外观(见图 1-e)，能学习语音指令，其内置传感器能感知食物投喂并对人们的触摸做出反应。一些设计原则已被整合到现有的类宠机器人中，例如机器人有拒绝或接受指令的自由，机器人的主人要帮助机器人成长并在此过程尽到对机器人的责任，使机器人能够形成对所有者的依赖(Kaplan, 2001)。类宠机器人交互研究能够分析出人们所需要的宠物行为模式，以便研究者设计并制造出更易被人们接受的类宠机器人。

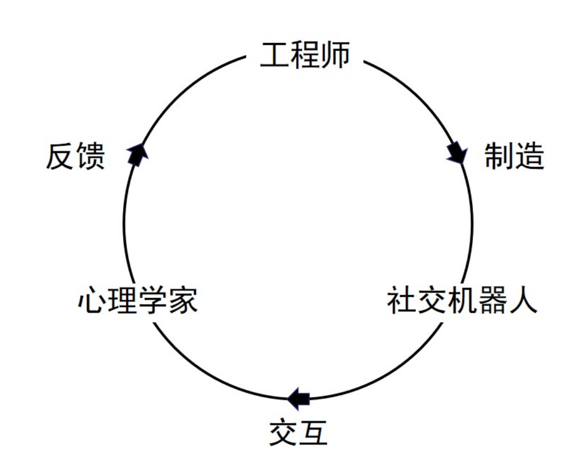


图 2 人与机器人交互研究环路

总体而言，人与机器人交互研究涉及多个学科领域，其中工程师负责制造社交机器人，而心理学家则通过使用这些机器人和真人进行交互来探索影响人与机器人二元关系的因素，并将这些研究成果反馈给工程师，使工程师进一步研发出交互能力更强的机器人(Broadbent, 2017)，由此形成如图 2 所示的研究环路。在这项跨学科的研究过程中，心理因素的重要性逐步增强。追求机器人的高复杂功能性并非人和机器人交互研究的首要目的，一种可行的研究策略是首先考虑机器人使用者的具体需求，再在此基础上发展机器人所因具备的功能(Vincent, Taipale, Sapio, Lugano, & Fortunati, 2015)，即研究环路中来自心理学家的反馈至关重要，这种以人为本的机器人设计理念让众多研究人员对人与机器人交互中所涉及的心理学问题展开了相关探索，以下部分对目前已有的研究主题进行分类梳理。

3.1 人与机器人交互中的信任

人对机器人的信任伴随机器人的发展逐渐受到研究者的关注。这种信任对人与机器人协作有效性有重要影响，在较高的信任水平下，人们可能会过分依赖并滥用机器人，而在较低的信任水平下，人们可能完全弃用机器人。考察人与机器人交互中人对机器人的信任水平是研究的重要主题之一。

目前人与机器人交互中信任问题的研究范式主要有以下四类：

(1)访谈。一项探索社交机器人接受度的研究中，研究者通过半结构化访谈来评估人们对机器人 Nabaztag (见图 1-f)的信任程度。Nabaztag 被带入参与者家中，并提供给参与者天气预报及身体锻炼建议等方面的信息。通过对访谈内容进行分析，研究者发现大部分参与者对机器人及其提供的信息表现出信任，但当机器人提供错误信息时，参与者对机器人的信任水平会下降(de Graaf, Allouch, & Klamer, 2015)。

(2)建模。有研究通过模拟战场环境，建立起人和机器人协作团队的评估系统，以此研究人和机器人协作任务中人对战术机器人的信任(Freedy, Devisser, Weltman, & Coeyman, 2008)。研究团队开发了多维协作表现模型，该模型涵盖了影响人与机器人协作信任的不同影响因素(如任务分配、机器人操控性等)，并通过仿真实验对这些因素的重要性进行了评估。

(3)信任博弈(trust game)。在实验室条件下，人们经常采用信任博弈任务来

考察人对机器人的信任。信任博弈(Berg, Dickhaut, & McCabe, 1995)的一般形式为：两位匿名玩家共同完成一项任务，双方分别拥有一定数额的金钱 S，要求一名玩家(委托者)把部分金钱 Y(O≤Y≤S)交予另一名玩家(受托者)，然后受托者获得 3Y 的金钱，并决定返还给信任者 X(0≤X≤3Y)的金钱，最终信任者收益为 S-Y+X，受托者收益为 S+3Y-X。在人和机器人的信任博弈中一般要求被试作为委托者与机器人完成该任务。采用这种范式，Haring 等人(2013)分析了人对类人机器人的信任。结果发现，性格越外向的被试越倾向于把金钱交予受托者(机器人)，即越外向的个体对机器人的信任度越高。

(4)囚徒困境博弈(prisoner's dilemma game, PDG)。除信任博弈外，也有研究者采取囚徒困境博弈评估人与机器人交互中的信任互惠。在此博弈任务中两名囚犯面临下述抉择，选择合作(彼此沉默)则两者都获得较短的刑期，选择背叛则背叛者会被释放而被背叛者会获得较长的刑期，有研究使用重复囚徒困境博弈来考察人与机器人交互中的信任互惠问题(Sandoval, Brandstetter, Obaid, & Bartneck, 2015)。实验中，研究者将互惠界定为对友好/敌对行为做出同类型的反馈，即机器人或真人代理在做出合作/背叛的选择后，被试也做出相同的选择。结果表明，相比机器人，被试更倾向于和真人进行合作，即被试更倾向于信任真人代理。

3.2 人与机器人交互中的共情

共情是指设身处地理解他人情绪和感觉的能力(Decety & Jackson, 2004)。共情通常被视为是社会交往中的重要能力之一，是社会合作及亲社会行为的基础。人与机器人之间的共情是影响人与机器人协作有效性的关键因素之一。实验室研究中，一般通过情景启动范式来研究人与机器人之间的共情，Leite 等人(2013)在研究中让机器人 iCat (见图 1-g)观察两位人类玩家间的象棋比赛，并对两位棋手的每一步走棋都做出评论。实验设定 iCat 对其中一方的评论充满共情，而对另一方的评论则比较中立，以此考察共情和中立两种评论反馈对人与机器人间共情的影响。通过对实验结束后被试所填写的开放式问卷进行分析，研究者发现那些受到机器人共情对待的被试认为机器人更加友好。还有研究者通过让被试观看机器宠物 Pleo 和真人受爱抚或虐待的视频来启动共情，并用功能磁共振成像(functional magnetic resonance imaging，fMRI)技术和量表相结合的方式对此进行研究，结果发现，当被试观看机器人或真人受爱抚的视频时，其脑神经激活模式是相同的。而相对于机器人受虐视频，被试观看真人受虐视频时会表现得更为痛苦(Rosenthal-von der Pütten et al., 2014)。另有研究通过给被试呈现被剪刀剪到的机械手图片来启动被试的共情，并用脑电图(electroencephalogram，EEG)描记法考察被试对真人和机器人的共情差异。结果表明被试对人和机器人的共情差异主要表现在自上而下加工的早期阶段，在其晚期阶段二者是相似的(Suzuki, Galli, Ikeda, Itakura, & Kitazaki, 2015)。

3.3 人与机器人交互中的社会距离

人际交往过程中，人与人之间通常会因亲近或疏远程度的不同而表现出不同的社会距离，这种距离既体现在生理方面也体现在心理方面。在人和机器人交互过程中，是否也存在这种现象？社会学家 Park (1942)提出了社会距离的概念并将其描述为：人与人之间在族裔、种族、宗教、职业等方面缺乏亲密关系的程度。有研究将人与机器人交互中的社会距离分为三个维度(权利距离、任务距离和空间距离)，并考察了社会距离变化时人对机器人的不同反应(Kim & Mutlu, 2014)。其中权力距离指的是人和机器人之间的上下级地位关系，任务距离指的是人和机器人执行任务时的结构(合作或竞争)，空间距离指的是人距离机器人的物理远近

程度。被试和机器人在不同社会距离条件下完成卡片匹配任务。结果表明，当被试面对下属机器人时，较远的空间距离会带给被试更好的体验。同时，被试对与机器人合作完成任务的体验评价高于和机器人对抗的体验评价。这些结果强调了机器人和人之间的社会距离及其行为表现的一致性，未来研究可进一步深入探索社会距离因素对人与机器人关系的影响。

3.4 对机器人的偏好

得益于机器人所带来的便利，人类在某些方面会表现出对机器人高于其他设备甚至真人的偏好。有研究对比了人对机器人和平板电脑的评价差异(Mann et al., 2015)，研究者让机器人 iRobiQ(见图 1-h)和平板电脑询问被试与健康相关的问题，并提供给被试完全一致的身体锻炼和放松练习建议。结果发现，相对于平板电脑，被试与机器人的互动更为积极（表现为言语和微笑次数的增加），并更愿意接受来自机器人的建议。访谈分析还表明，被试对机器人以及人和机器人关系的评价更为积极，与机器人再次互动的意愿也更高。另有研究揭示了在向导指引方面，人对机器人高于真人的偏好(Satake et al., 2015)。研究者对在商场中使用向导机器人的顾客进行了回访，结果表明相对于真人向导，顾客更倾向由机器人向导提供诸如发放传单、线路指引等方面的服务。这种倾向性并非因为机器人所提供的服务质量优于真人，而是因为顾客对机器人会表现出好奇心。同时，顾客向机器人提出服务请求时不会像面对真人时表现出尴尬和羞涩。

总体而言，在这些人与机器人交互的研究主题中，研究者都试图从心理学角度对人与机器人二元关系的建立和发展进行探索。人与机器人交互研究离不开相关理论的支持，以下部分对人与机器人交互研究的主要理论进行简要概括。

4 人与机器人交互研究的主要理论

4.1 恐怖谷理论

Mori(1970)是最早研究人和机器人交互的科学家之一，他于 20 世纪 70 年代提出，机器人的拟人程度和人面对它们时内心舒适感之间的关系会随其拟人程度的增加而发生如图 3 所示的变化：机器人的外表越接近人类，人内心的舒适感越高，但当机器人的外表极度接近真人又不完全和真人一致时，舒适感会出现陡降，这种现象被称为恐怖谷(Mori, 1970; 邓卫斌，于国龙，2016)。Mori 还阐述了运动对恐怖谷的影响，即机器人的运动方式也会改变人面对机器人的舒适感。为使人们在和机器人交互的过程中感到自然舒适，对于类人社交机器人而言，其拟人程度应尽可能高，但需注意避免因拟人程度过高而陷入“恐怖谷”。有研究者对恐怖谷理论进行了解释：当被试观察到机器人的行为与预期不相符时，大脑中便会产生错误反馈信号。换言之，若机器人长相接近真人，却以一种怪异的方式运动，即当机器人“形”似人而“行”不似人时，便会出现对直觉期望的冲突，进而让人产生恐惧心理(MacDorman & Chattopadhyay, 2016)。尽管存在类似研究进一步支持了恐怖理论并对其做出了其他合理解释，但针对该理论仍存在较大争议，原因主要包括三点：首先，舒适感作为因变量缺乏统一的定义和有效的测量；第二，拟人程度作为自变量受多重因素影响，难以明确界定并系统操纵；第三，缺乏明确的数学模型拟合恐怖谷理论所描绘的曲线(Wang, Lilienfeld, & Rochat, 2015)。理解恐怖谷理论的本质和恐怖感觉产生的边界条件可有效引导社交机器人的设计，避免因机器人的拟人程度过高而对人和机器人的交互产生负面影响。

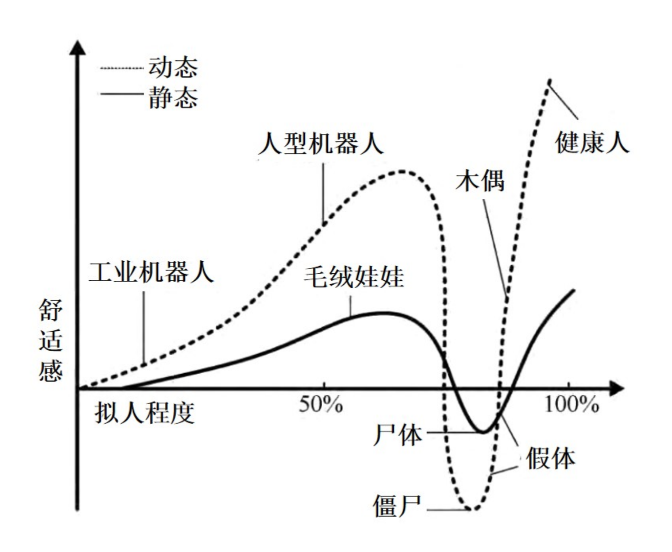


图 3 恐怖谷

4.2 媒体等同(media equation)理论

研究者发现，当人们和一些无生命的物体交互时，尽管他们知道这些物体是无生命的，但还是会无意识地赋予其人际交互中的社会规范，这种理论被称为媒体等同(Reeves & Nass, 1996)。有实验证据支持了这一理论，实验者使用电脑向被试提供信息，结果发现被试认为具有女性声音的电脑更有爱心，具有男性声音的电脑更富有学识(Nass, Moon, & Green, 1997)。这与日常生活中人们对男女两性的刻板印象一致，即女性通常被认为具有更好的关怀能力，而男性具有更强的主观能动性(Huddy & Terkildsen, 1993)。不仅对电脑，当人们和机器人进行交互时，也会无意识地把社会规范赋予机器人。在一所大学中进行的实验表明，接近 50%的被试和机器人接待员进行深入交互之前，会像和真人进行交互时一样，以某种方式向机器人打招呼，这种问候的存在预示着被试与机器人随后进行的交互将更社会化并具有礼貌性(Min, Kiesler, & Forlizzi, 2010)。媒体等同的理论还体现在种族偏见方面，外观功能一致的两个机器人，德国国籍的被试会认为具有德国名字且制造于德国的机器人比另一个被描述为具有土耳其名字且制造于土耳其的机器人具有更好的设计及性能，并能让人感到更加亲密(Eyssel & Kuchenbrandt, 2012)。这些结果都表明在人和机器人的交互过程中，人们不仅仅把机器人当做工具使用，还赋予其一定的社会属性，人与机器人二元关系将随着这些属性的加入而进一步复杂化。

4.3 心智感知(mind perception)理论

机器人是否具有心智？最早关注这个问题的研究者 Alan Turing 在其著名的图灵测试中设置下述场景：测试者分别向真人和机器进行提问，如果测试者无法根据答案来区分被测试者是人还是机器，那么这台机器就被认为具有人类智能(Turing, 1950)。一般认为人的心智由两部分组成，即代理的能力(如自我控制，情绪识别，计划，交流，思考，道德品行等)和体验的能力(如，喜、怒、忧、悲、恐，惊等)(Gray, Gray, & Wegner, 2007)。所谓心智感知，指的是人类在与机器人交互过程中会感知到机器人不同程度的心智，Gray 等人(2007)的研究阐明了这一理论。他们研究发现，机器人被认为具有较少的体验能力，但具有中等程度的代理能力，这表明大多数人在机器人身上感受到了心智。即使目前的科技水平还无法让机器人像真人一样具有自我意识，至少在与其进行交互的个体看来，它们在

某种程度上都具有心智和思想。

5.伦理思考及研究展望

美国著名科幻小说家 Isaac Asimov (1950)在其著作中提到了机器人三大定律，即第一定律：机器人不得伤害人类个体，或者目睹人类个体遭遇危险而袖手不管；第二定律：机器人必须服从人给予它的命令，当该命令与第一定律冲突时例外；第三定律：机器人在不违反第一、第二定律的情况下要尽可能保护自己的生存。这三条定律源于科幻小说，但也有其现实意义，机器人是人类智慧的产物，那么机器人是否像人类一样拥有自身权利？一项关于人们对向导机器人态度的研究表明，儿童和父母在一起时，他们对机器人的态度比较友好。当缺少成人看管时，儿童可能会聚集起来虐待机器人，如封锁机器人的道路，辱骂踢打机器人并忽视机器人劝停的礼貌请求等(Brscić, Kidokoro, Suehiro, & Kanda, 2015)。设计者应当在机器人的外观和行为方面做出合理的设计来规避机器人的受虐风险。此外，这一领域的研究成果对人类了解自身行为及降低动物受虐风险也有参考价值。人类要求机器人对其绝对忠诚和服从的同时，是否也应该考虑人类对机器人的责任，如尊重机器人的工作、合理回收报废机器人等等，这些人和机器人交互中的道德问题值得研究者深思并将之运用于社交机器人的设计研发当中。

囿于当前的科技水平，现有的人与机器人交互研究范式中，虽然被试被告知他们是在和机器人进行交互，但实际上这是一种伪交互，与之交互的机器人大多由真人操纵。类似于中国传统文化中的木偶戏，实验者在幕后控制了机器人的动作和声音。这种操控难免会让实验的参与者感到机器人的反应生硬不自然。鉴于此，人类在与机器人的交互过程中，可能无法拥有和真人交互一样的情感体验。制造机器人需要克服诸多工程学上的难题，让机器人具有类人的情感、认知和行为更是严峻的挑战。面对这种挑战，人们需要打破由工程师和计算机专家主导人工智能(artificial intelligence)开发的局面，让包括心理学在内的更多学科融入到该领域。现有的人与机器人交互研究主题需要进一步深入和扩展。

5.1 信任主题

信任作为重要的社会信号机制之一, 通过降低社会交易成本, 易化合作行为成为建立良好社会关系的基石。人际互动中的信任研究通常采用信任博弈(TG)范式并利用脑成像技术探究信任行为背后的认知神经机制。王益文等研究者(2015)利用重复信任博弈任务(repeated trust game, rTG)探索了个体在信任互动情境下大脑活动变化的时间动态特征。结果发现，在决策阶段，不信任选择比信任选择诱发了更正的 P2 成分(150~250 ms)，反馈阶段中损失反馈比获益反馈诱发了更负的 FRN 成分(200~300 ms), 而获益反馈比损失反馈诱发了更短的 P300 潜伏期。行为结果还表明个体选择信任的比例显著高于几率水平。在人和机器人交互的信任中是否存在同样的大脑活动特征及行为结果有待进一步考证。目前鲜有研究者在人与机器人交互的信任研究中结合功能性磁共振成像技术(fMRI)以及事件相关电位(event-related potentials, ERPs)等技术对个体与真人和机器人交互中信任的认知神经机制的异同进行分析。未来人与机器人交互的信任研究可充分利用人际交互信任研究中已采用的 fMRI，EEG，ERPs 等技术，设置符合机器人实际应用的交互场景来探索人和机器人交互信任的认知神经机制。进而通过调整机器人的拟人程度等可能影响人与机器人信任的因素，提升人们对机器人的信任水平，使人类和机器人在适度的信任水平下进行交互以取得良好的交互效果。

5.2 共情主题

共情可以对可接受的社会行为的形成和发展进行调节，它包含两个方面的因素，其一是分享他人情感状态的情绪反应，二是站在他人角度看待问题的认知因素(Decety & Jackson, 2004)。机器人表现出可接受的社会行为，融入人类社会离不开人们对其所产生的情感影响的研究。目前已有研究通过 fMRI 技术和 EEG 技术考察了个体与真人或机器人交互时的共情差异(Leite et al., 2013; Suzuki, Galli, Ikeda, Itakura, & Kitazaki, 2015)，但这些研究中所使用的共情情景启动范式与机器人在日常生活中的实际应用存在一定的差距，机器人评论棋手的走棋或被剪刀剪到手指并非经常出现在其应用的主要场景。未来研究需着眼于机器人的实际应用，在医院陪护、学生教育、商场向导等使用情境下，探索人和机器人交互的共情问题，以发掘更适合机器人在上述情境下应用的设计。

5.3 群体身份主题

在日常生活中，通常存在个体凭借某些标签将互动对象进行组内/组外成员的清晰区分，这种区分可使个体感知到互动对象是否和自己同属某一社会群体，进而影响个体互动时的心理加工与行为决策，这种理论被称为社会认同理论(Tajfel, 1978)，相比外群体成员，个体更愿意与内群体成员分享成果、共同承担消极结果，并且对内群体成员的情绪共情能力也显著高于对外群体成员的共情能力，在内群体成员中会存在更多的合作、互惠行为，更容易达成协议也更注重个体间的公平。在人际互动中，有研究者利用最后通牒任务(ultimatum game, UG)考察了群体身份对 UG 博弈中反应者公平关注的影响及其动态时间过程，结果发现在群体互动情境下，互动成员的群体身份能够影响个体的早期注意资源分配和公平关注加工(王益文等，2014)。在人和机器人互动的过程中，机器人是否会被与之互动的个体归为外群，机器人的群体身份(组内/组外)对个体心理加工和行为决策的影响及其动态过程是否和与真人互动时一致等相关问题值得进一步探究，相关研究结果可帮助工程师制造出更符合人类群体的机器人。

总体而言，人和机器人交互中的信任、共情等研究主题还可进一步深入探索。人与机器人交互研究可参照现存的人际交互研究主题，利用人际交互研究范式，如信任博弈(trust game)、懦夫博弈(chicken game)、最后通牒博弈(ultimatum game)等进行扩展和丰富，以探究人和机器人交互与人际交互的异同，并揭示交互过程的心理加工机制及认知神经机制。在此基础上进一步考察人与机器人二元关系的建立、变化及其影响因素，为机器人更好地融入人类社会及制造更适宜人类社会的机器人提供有效引导。

现有研究大多关注机器人的机械工程和软件工程层面，聚焦于人和机器人交互本身的研究还较为匮乏。伴随人工智能的快速发展，机器人的功能愈发强大，如何让高度智能化的机器人融入人类社会，而不仅仅沦为炫技的展品，将是未来研究需要探索和解决的问题。换言之，人工智能领域需要心理学的融入以形成机器人心理学或人工智能心理学这些新兴交叉学科，来帮助人工智能及其对应产品以更易被人类接受的方式进行发展。

机器人护理、无人机快递等人工智能技术的出现，让我们看到了未来人工智能取代人类完成一些体力劳动和脑力劳动的可能。尽管人工智能能够凭借其优化的算法及高速的计算水平在某些工作领域（如车辆装配、出纳收银等）具备优于人类的表现，但在与人类交互方面，它也许还无法自然且饱含情感地和人类进行沟通。同样，在一些需要创意性思维和启发式思考的领域，人工智能可能还未能达到人类的水平。在人工智能技术进一步扩展和优化的过程中，人类如何与人工

智能共同生存值得深思。心理学的融入恰好可帮助人们更好地理解和解决人和人工智能的共生问题。已问世的人工智能技术如何被人们接受并推广使用，仍处于研发阶段的人工智能技术如何融入人类的习俗、社会规范等相关探索正是心理学融入人工智能领域需要发展的方向。心理学和人工智能的交叉将使人类更好地迎接人工智能时代所需面对的挑战，同时，这两个领域的交叉也将使心理学在人工智能改变世界的过程中发挥出应有的功能。科学技术的发展应做到以人为本，以人类的切身需求为中心，无论是在人工智能还是机器人的研发过程中，都不能仅仅局限于对其实用功能的扩展，还应充分考虑到心理因素的重要性，使这些产品更好地融入人类社会并造福于人。在人工智能认知结构设计，人工智能学习模仿人类情感和行为，人工智能辅助疗法等方面，心理学都将起到至关重要的作用。

致谢 特别感谢宁波大学尹军博士对本研究提供的意见和帮助。

参 考 文 献

[1]邓卫斌,于国龙. (2016). 社交机器人发展现状及关键技术研究. *科学技术与工程, 16*(12), 163–170.

[2]王益文，张振，张蔚，黄亮，郭丰波，原胜. (2014). 群体身份调节最后通牒博弈的公平关注. *心理学报, 46*(12), 1850–1859.

[3]王益文，张振，原胜，郭丰波，何少颖，敬一鸣. (2015). 重复信任博弈的决策过程与结果评价. *心理学报, 47*(8), 1028-1038.

[4]Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142.

[5]Broadbent, E., Tamagawa, R., Kerse, N., & Knock, B. (2009). *Retirement home staff and residents' preferences for healthcare robots.* Paper presented at the Ro-Man 2009 - the IEEE International Symposium on Robot and Human Interactive Communicatio (pp. 645–650). Piscataway, NJ: IEEE.

[6]Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, 68, 627-652.

[7]Brscić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015). *Escaping from children's abuse of social robots*. Paper presented at the Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction – Hri (pp. 59–66). New York: ACM.

[8]Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Journal of Educational Technology & Society*, 13(2), 13-24.

[9]de Graaf, M. M. A., Allouch, S. B., & Klamer, T. (2015). Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior, 43*, 1–14.

[10]Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71–100.

[11]Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disords, 6*(1), 249–262.

[12]Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: anthropomorphism as a function of robot group membership. *British Journal of Social Psychology, 51*(4), 724–731.

[13]Frédéric Kaplan. (2001). *Artificial attachment: will a robot ever pass ainsworth's strange situation test*. Paper presented at the Proceedings of second IEEE-RAS International Conference on Humanoid Robots (pp. 125–132). Piscataway, NJ: IEEE.

[14]Freedy, A., Devisser, E., Weltman, G., & Coeyman, N. (2008). *Measurement of trust in human-robot collaboration.* Paper presented at the International Symposium on Collaborative Technologies and Systems (pp. 106–114). Orlando, FL: IEEE.

[15]Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619.

[16]Han, J., Park, I. W., & Park, M. (2015). *Outreach education utilizing humanoid type agent robots*. Paper presented at the Proceedings of the 3rd International Conference on Human-Agent Interaction (pp. 221–222). New York: ACM.

[17]Haring, K. S., Matsumoto, Y., Watanabe, K., Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). *How do people perceive and trust a lifelike robot?* Paper presented at the International Conference on Intelligent Automation and Robotics (vol 1, pp. 425 430). San Francisco, USA.

[18]Isaac Asimov. (1950). *I, robot*. United States: Gnome Press.

[19]Ishiguro, H., & Nishio, S. (2007). Building artificial humans to understand humans. *Journal of Artificial Organs, 10*(3), 133–142.

[20] Kim, Y., & Mutlu, B. (2014). How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies, 72*(12), 783–795.

[21] Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human–robot relations. *International Journal of Human-Computer Studies, 71*(3), 250–260.

[22] MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition, 146*, 190–205.

[23] Mann, J. A., MacDonald, B. A., Kuo, I. H., Li, X., & Broadbent, E. (2015). People respond better to robots than computer tablets delivering healthcare instructions. *Computers in Human Behavior, 43*, 112–117.

[24] Min, K. L., Kiesler, S., & Forlizzi, J. (2010). *Receptionist or information kiosk:how do people talk with a robot?* Paper presented at the ACM Conference on Computer Supported Cooperative Work (pp. 46–47). New York: ACM.

[25] Mori. (1970). The uncanny valley. *Energy. 7*, 33–35.

[26] Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., & Dong, J. (2013). A review of the applicability of robots in education. *Technology for Education and Learning, 1*, 209–215.

[27] Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27*(10), 864–876.

[28] Park, R. E. (1924). The concept of social distance: as applied to the study of racial relations. *Journal of Applied Sociology. 8*, 339–344

[29] Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). *What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic roBOT (ABOT) database*. Paper presented at the Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 59–66). New York: ACM.

[30] Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York, US: Cambridge University Press.

[31] Robinson, H., Macdonald, B., Kerse, N., & Broadbent, E. (2013). The psychosocial effects of a companion robot: a randomized controlled trial. *Journal of American Medical Directors Association, 14*(9), 661–667.

[32] Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., & Maderwald, S., et al. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior, 33*, 201–212.

[33] Rosheim, M. (2006). *Leonardo΄s lost robots*. Berlin: Springer.

[34] Sabelli, A. M., & Kanda, T. (2015). Robovie as a mascot: A qualitative study for long-term presence of robots in a shopping mall. *International Journal of Social Robotics, 8*(2), 211–221.

[35] Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2015). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics, 8*(2), 303–317.

[36] Satake, S., Hayashi, K., Nakatani, K., & Kanda, T. (2015). *Field trial of an information-providing robot in a shopping mall.* Paper presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 1832–1839). Piscataway, NJ: IEEE.

[37] Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, 5.

[38] Tajfel H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London: Academic Press.

[39] Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

[40] Vallèsperis, N., Angulo, C., & Domènech, M. (2018). Children's imaginaries of human-robot interaction in healthcare. *International Journal of Environmental Research & Public Health*, 15(5), 970.
[41] Vincent, J., Taipale, S., Sapio, B., Lugano, G., & Fortunati, L. (2015). Social robots from a human perspective. Berlin: Springer.
[42] Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: existence and explanations. *Review of General Psychology*, 19(4), 393-407.
[43] Wang, X., & Krumhuber, E. G. (2018). Mind perception of robots varies with their economic versus social function. *Frontiers in Psychology, 9*(1230).

Current status and Perspectives of Research for Human-Robot Interaction——Elementary Discussion on the Interdiscipline of Psychology and Artificial Intelligence

Deng Zhaoxin[1,2] Fu Chao[1,2] Yang Xue[1,2] Wang Yiwen[1,2]

[1](Institute of Psychological and Cognitive Sciences, Fuzhou University, Fuzhou 350116, China)

[2](Center for China Social Trust Studies, Fuzhou University, Fuzhou 350116, China)

Abstract

Robots are slowly being incorporated in our society, previous studies mainly focused on software programming or mechanical manufacturing of robots, there's limited research conducted with the view to highlight the importance of psychology in human-robot interaction, especially in the establishment, development and change of the binary relationship between human and robots. Since social robots are more social compared to industrial robots, human-robot interaction particularly refers to the interaction between human and social robots. On the basis of introducing the main applications of social robots, this paper classifies and concludes the related themes of research for human-robot interaction including trust, empathy and social distance, and expounds the theory of uncanny valley, media equation and mind perception involved. The possible ethical issues caused by social robots are discussed as well. This paper also targetedly penetrates into and extends the future research themes in human-robot interaction, including but not limited to trust and empathy. Psychology can inform us about human emotion, cognition and behavior, it also helps to establish and shape a good binary relationship between human and robots.

Key words: social robots; human-robot interaction; artificial intelligence psychology